

CERGE - EI

Centre for Economic Research and Graduate Education –
Economics Institute
Charles University



Effects of voters' privacy on the public choice: How the microtargeted political advertisement affects the election results and voters' behaviour

Vít Illichmann

Master Thesis

Prague, July 2023

Author: Bc. Vít Illichmann

Supervisor: Ole Jann, Ph.D. ¹

¹Centre for Economic Research and Graduate Education – Economics Institute

I declare that I carried out this master's thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

In date

signature of the author

Acknowledgments

I would like to express my gratitude to my supervisor, Ole Jann, Ph.D., who provided a valuable expertise and support throughout the whole process of writing the thesis.

Master's Thesis Proposal

Author's name: Bc. Vít Illichmann **e-mail:** Vit.Illichmann@cerge-ei.cz

Supervisor's name: Ole Jann, Ph.D. **e-mail:** Ole.Jann@cerge-ei.cz

Proposed topic:

Effects of Voters' Privacy on the Public Choice: How the microtargeted political advertisement affects the election results and voters' behaviour.

The preliminary scope of work:

Research question and motivation

The goal of the thesis is to review the latest empirical and theoretical research regarding online microtargeting practices in political campaigning and their effect on the voter's behaviour. The primary aim concerning the proposed topic is to assess if the lack of voter privacy in the online environment allows political candidates to influence voters' decisions via microtargeted political advertisements. Given the heterogeneity of the theoretical results in the related research, the author aims to conduct a meta-analysis of empirical results on the topic and identify a set of results that appear to be robust across multiple studies with different analytical approaches and confront them with conclusions stemming from the theory. Based on the set of stylized facts identified within the body of empirical work, the author wants to propose a computational model and study the effects of the voter-candidate information asymmetry on the election outcome via means of simulation.

Methodology

The author views the interpretation of theoretical results within the context of recent empirical research as a valuable and necessary contribution. Given the fact that theoretical models in privacy economics yield results that are often highly context-sensitive, such contrast is necessary for further meaningful discussion of policies regarding online privacy regulations. Further, the accompanying model should supply the reader with intuition about the effects of informational asymmetry caused by the microtargeting of the voter on the results of public choice.

Outline

1. Introduction
2. Literature review
3. Review of theoretical models
4. Review of empirical research
5. Comparison of results
6. Model setup and analysis
7. Summary of results and policy implication
8. References

Bibliography

1. Acquisti, A., Taylor, C., & Wagman, L. (2016). The Economics of Privacy. *Journal of Economic Literature*, 54(2), 442–492. <https://doi.org/10.1257/jel.54.2.442>
2. Burkell, J., & Regan, P. M. (2019). Voter preferences, voter manipulation, voter analytics: policy options for less surveillance and more autonomy. *Internet Policy Review*, 8(4). <https://doi.org/10.14763/2019.4.1438>
3. Hoffmann, F., Inderst, R., & Ottaviani, M. (2019). Persuasion Through Selective Disclosure: Implications for Marketing, Campaigning, and Privacy Regulation. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3811939>
4. Matthes, J., Hirsch, M., Stubenvoll, M., Binder, A., Kruike-meier, S., Lecheler, S., & Otto, L. (2022). Understanding the democratic role of perceived online political micro-targeting: longitudinal effects on trust in democracy and political interest. *Journal of Information Technology and Politics*, 1–14. <https://doi.org/10.1080/19331681.2021.2016542>
5. Lavigne, M. (2020). Strengthening ties: The influence of microtargeting on partisan attitudes and the vote. *Party Politics*, 27(5), 965–976. <https://doi.org/10.1177/1354068820918387>
6. Prummer, A. (2020). Micro-targeting and polarization. *Journal of Public Economics*, 188, 104210. <https://doi.org/10.1016/j.jpubeco.2020.104210>
7. Madsen, J. K. (2019). The Psychology of Micro-Targeted Election Campaigns. <https://doi.org/10.1007/978-3-030-22145-4>
8. Witzleb, N., & Paterson, M. (2021). Micro-targeting in Political Campaigns: Political Promise and Democratic Risk. Data-Driven Personalisation in Markets, *Politics and Law*, 223–240. <https://doi.org/10.1017/9781108891325.017>
9. Zarouali, B., Dobber, T., De Pauw, G., & de Vreese, C. (2020). Using a Personality-Profiling Algorithm to Investigate Political Microtargeting: Assessing the Persuasion Effects of Personality-Tailored Ads on Social Media. *Communication Research*, <https://doi.org/10.1177/0093650220961965>

Contents

1	Introduction	7
2	Literature review	14
2.1	Micro-targeting and polarization of policy	15
2.2	Micro-targeting as a strategic advantage	19
2.3	Micro-targeting in empirical research	23
2.4	Discussion	26
3	Model	32
3.1	Introduction to the model	32
3.2	Primitives of the model	35
3.2.1	Voters	36
3.2.2	Candidates	41
3.3	Solution of the model	44
3.3.1	Optimal signalling actions of the voters	44
3.3.2	Posterior beliefs of the candidates	46
3.3.3	Posterior beliefs of the voters	46
3.3.4	Optimal campaign strategy of the candidate	47
3.3.5	A moderate candidate running against an extreme candidate	48
3.4	Discussion	52
4	Conclusion	57
5	References	62
5.1	Academic references	62
5.2	Non-academic references	65

Abstract

The thesis analyses the phenomenon of targeted political advertising and its influence on policy. In addition, it is concerned with incentives for candidates to breach the privacy of voters. To aid my analysis, I build a simple voting model with imperfect information, where extreme and moderate candidates compete to maximise their proportional electorate support. I show that the information aggregation process on the side of the candidate significantly changes voting outcomes, and insufficient privacy protection increases equilibrium support for the extreme candidate. Further, I show that the extreme candidate has a higher incentive to breach the privacy of the voters in comparison with his moderate opponent. The result of my analysis might have policy implications regarding the regulation of targeted political advertisements on social media. My results rely on the assumption of behavioural conformity of political discourse on social media, which is an empirically documented phenomenon. My thesis adds to a sparse strain of theoretical literature on targeted political advertising.

Key words: Targeted political advertising, Online privacy protection, Policy polarisation, Social media, Micro-targeting

Abstrakt

Tato diplomová práce analyzuje fenomén cílených politických kampaní a jeho vliv na veřejnou politiku a zabývá se existencí incetivů pro politické kandidáty k zásahu do soukromí voličů. Pro účely analýzy využívá matematický model hlasování s nekompletní informací, ve kterém kandidáti usilují o maximalizaci proporcionalního podílu získaných hlasů. Výsledky analýzy ukazují, že proces shromažďování informací ze strany politických kandidátů může ovlivnit výsledky hlasování. Dále demonstrují, že kandidáti zastávající extrémní pozice mají větší motivaci využívat v rámci svých kampaní soukromé informace voličů ve srovnání s kandidáty zastávajícími centrální pozici. Výsledky mé analýzy mohou přispět do k diskusi o regulaci cílené politické reklamy na sociálních mediích. Prezentovaný model předpokládá existenci konformity ve vyjadřování politických názorů na sociálních mediích, tento jev je dokumentovaný empirickými poznatky. Diplomová práce rozšiřuje teoretickou literaturu týkající se cílené politické reklamy a jejího vlivu na demokratický proces.

Klíčová slova: Cílená politická reklama, Ochrana soukromí online, Polarizace veřejných politik, Sociální media, Micro-targeting

1. Introduction

This thesis aims to explore the systemic effects of targeted political advertising. In particular, I focus on incentives for candidates to breach voters' privacy and track their behaviour across social media platforms. My contributions to the existing body of literature are the following: Firstly, I provide a review of the empirical and theoretical works and contrast the results. Secondly, I propose a simple theoretical model which takes into account the information acquisition process on the side of the political candidates, showing that incentives to breach privacy may vary depending on the candidate's political position with respect to the mean political preferences of the electorate; more specifically, I show that more extreme candidates can have different payoffs from the use of private information than moderate politicians. My results thus potentially add to the ongoing debate about the regulation of political advertising on social media.

In comparison with the original thesis proposal, I put less emphasis on the literature review in favour of the model. Additionally, I decided for the analytical solution of the proposed model as opposed to the originally intended computational simulation, as the analytical results are more transparent.

Mapping the preferences of the electorate is essential in order to ensure success in elections and acquire political influence. In the recent past, data about the public's political attitudes were obtained via systematic opinion surveys and campaigning conducted either through direct contact with the broader public or by disseminating political advertisements through mass media outlets. With the increasing popularity of online outlets, in particular social media platforms, campaigning conducted by traditional means prevailed;

however, a substantial portion of public communication with voters moved online. Considering the immensely broader reach of online platforms, comparatively lower costs and a higher level of audience engagement, this is a perfectly natural development. Nowadays, it is common for high-profile political figures to maintain social media accounts across multiple platforms operated by an experienced PR team. Social media posts by high-profile politicians tend to have a long-lasting impact on their public image as well as steer the flow of online political debate. Anecdotal evidence documenting such behaviour is ubiquitous, and it is a well-documented phenomenon in academic research as well (e.g. Silva, Proksch, (2021)). Unlike the sterile public discussions in mass media, which allow direct interaction only with a limited audience, online discourse allows arbitrary users to interact with each post. Consequently, voters can take publicly observable actions to express their approval by sharing, liking or otherwise engaging with political content on their feeds. This mode of user engagement can be observed by the candidates as well as by the peers of each voter connected to him in the social network, i.e. by his friends, coworkers, schoolmates or family members. Assuming each user enjoys promoting his views with an intensity that depends on some exogenous psychological factors, there is an inherent incentive for people to engage with content that aligns with their views. However, each user has to consider the implications of explicit or implicit public expression since there might be adverse effects associated with openly expressing controversial views regarding political topics. Consequently, voters with fringe political beliefs are less likely to take part in online public discourse because of the increased risk of harsh treatment by their peers. An example of a scenario where fear of harsh treatment leads to secondhand conformity in the online debate can be seen in Jann and Schottmüller

(2019). The authors show that in the case where information shared online can be used to guess some unobservable characteristics of the agent, the agents with controversial views are more likely to keep their opinions private to avoid statistical discrimination. Empirical evidence of self-censorship in online communication networks is also provided by Madsen and Verhoeven (2016), who study communication patterns of employees on internal social media in a Danish bank. Another interesting showcase of such behaviour is the study by Burnett, Knighton and Wilson (2022) which directly dissected the motives leading people to refrain from sharing their political opinions, showing that the leading cause of political self-censorship is the fear of social isolation.

The rapid development of communication technology not only increased the share of social media in the traditional mode of a political campaign where one candidate communicates a uniform message to a large-scale audience but also opened the door for a new mode of communication that allows marketers to diversify messaging across audiences based on differential preferences of smaller online communities or even individuals. Such a mode of mass online communication which is centred around profiling individual preferences based on a collection of user data is usually referred to as micro-targeting or targeted advertising within related literature. The term micro-targeting naturally encompasses a wide array of advertising practices which may substantially differ from each other. For the purpose of this thesis, I consider micro-targeting to be any form of online advertising which requires some form of individual-level data.

Targeted advertising on social media is a common practice in commercial contexts; the global market for social media advertising is projected to reach a value of 207 billion USD in 2023 (Statista,

2023). In order to profile users for commercial purposes, marketers use data purchased from online service providers. The level of detail contained in such data varies greatly depending on the data provider, commercial field and the associated legal context. To gain insight into the structure of the data collection schemes, see Ullah, Boreli and Kanhere (2022); the discussion concerning technical aspects of data collection and privacy protection solutions is beyond the scope of this thesis.

Generally speaking, collected data include metadata from visited websites, search queries, metadata about sent messages and time logs about active engagement with the online content. To combat the risk of privacy infringement, most advanced countries already have a complex legal framework in place that dictates general rules of user data protection, such as the GDPR regulation active within the EU legal area.

But then again, there is a massive informational asymmetry between regulators and technology companies, allowing service providers to take advantage of legal loopholes. In order to provide a readily understandable example of such regulatory leakage, I can present the use of Google fonts on websites. Every website on the internet needs a source of typography in order to display static text. Most popular websites use a Google font API to obtain their typography; consequently, the IP of each user visiting an arbitrary website with Google fonts API gets logged in directly by Google, even in case the website itself does not collect any personal data. The purpose of this short digression was to showcase that although privacy protection is already in place, it is hard to monitor and enforce from a practical point of view. Due to the high level of interconnectedness between various online services, it is increasingly hard to maintain online privacy as a user because the data mapping individual activity can

be collected on various levels. For example, the same information can be logged in by the active app as well as the owner of the website the user is interacting with. Further, there is evidence that certain applications can collect data about traffic in other unrelated tools. Consequently, if the user wishes to protect his privacy, the monitoring costs on his side can be quite high. Current regulation typically allows users to opt out of the service in case he disagrees with the privacy conditions. Consequently, under the current framework, most of the responsibility rests on the individual consumer. Empirical evidence suggests that people typically underestimate the risk associated with leakages of private data (Uno, Sonoda, Bessho, 2021). Thus, although the regulatory framework is very strict in some aspects, it does not take into account the incomplete knowledge on the side of the consumer. There is thus an obvious disparity between regulation applied in the technological sector and other industries which deal with a large informational asymmetry between the customer and service provider. Similar asymmetry is present, for example, in the banking industry; however, there, the regulatory requirements are largely skewed in favour of the consumer to control for undesirable outcomes.

In the political context, the notion of micro-targeting was popularised mainly by the Cambridge Analytica scandal surrounding the Brexit vote and the 2016 US presidential elections, sparking a heated public debate regarding social media companies' careless treatment of users' data. While there are undisputable ethical concerns associated with the use of individual-level data for the purposes of tailored political advertising, the strategic implications of employing targeted political advertising in election campaigns remain unclear. Much of the scientific literature concerned with micro-targeting in politics focuses on the exploitation of preexisting biases of voters and other

behavioural imperfections. Madsen (2019) provides a detailed discussion of the psychology behind modern political campaigns; much of the success of data-driven campaigns seems to be attributable to subjective reasoning flaws and differences in prior beliefs of voters with different socioeconomic backgrounds. It is intuitive that in the case when only one of the candidates uses targeted advertising, he has a competitive advantage solely due to the fact he is better informed in comparison with his opponents. For example, Hoffmann, Inderst and Ottaviani (2013) extend their model of commercial micro-targeting to political campaigns; the results show that voters are worse off in the equilibrium if only one of the candidates chooses to target.

However, if all candidates have access to the same technology and data, the question of whether one of them gains a competitive advantage is not trivial. The model by Titova (2023) suggests that micro-targeting is advantageous for newcomer candidates who challenge the status quo policy. The result relies on the fact that the policy platform of the status quo candidate is a priori known, while there is a remaining uncertainty about the platform of the challenger. Consequently, the challenging candidate can take advantage of the existing ambiguity and send private messages to sway voters on both sides of the political spectrum to support his platform. The long-run effects of micro-targeting on election outcomes are discussed by Prummer (2020), who shows that targeted political advertising, together with the increasing fragmentation of online media, adds to the long-run polarization of policy. Both models, that is, Prummer (2020) and Titova (2023), work under the assumption that the candidates know the political preferences of ex-ante individual voters; hence no results regarding the privacy of the voters arise, and there are no costs of information acquisition on the side

of the candidate.

In this thesis, I present a model of election competition in which candidates initially know only the population distribution of political preferences. To infer individual-level preferences, the candidate can either monitor the publicly observable reactions of voters towards presented policy proposals or enter a contract with a service provider and purchase the private information of individual voters. The voters are modelled as conformists and take into account the preferences of their peers when engaging with political content; consequently, the voters with fringe political views are less likely to partake in the public discourse. My aim is to show if the online public debate is inherently conformist; the information acquisition process on the candidate's side is not sterile but may influence the election outcomes as well. The results of the model suggest that although all candidates are incentivised to micro-target their supporters, regardless of their position on the political spectrum, extremists have a higher incentive to use private data to identify their voter base.

The rest of this thesis is structured as follows, the next chapter 2 gives an overview of academic literature related to political campaigns and provides a detailed review of two theoretical models that directly cover the topic of political micro-targeting. The subsequent section 3.1 introduces the general structure of the model. Section outlines in detail the description of preferences, strategies and actions of voters and candidates. Section 3.3 gives the key results of the model and continues with a discussion in section 3.4. The conclusion follows.

2. Literature review

In this section, I review the selected academic works relevant to the targeted political advertisement. There are several strains of theoretical literature which are relevant with respect to this topic. For example, the literature on strategic communication on networks provides certain insight into the dynamics of social media campaigns (e.g. Galeotti, Ghiglino, Squintani, (2013)).

However, rigorous theoretical models addressing targeted political advertising as such are sparse. Consequently, I choose the following approach; I give a detailed review of articles by Prummer (2019) and Titova (2023). Prummer (2019) is concerned with the dynamic effects of micro-targeting on policy in the context of increasingly fragmented online media. The main focus of Titova (2023) is a distinction between private and public communication in political campaigns. I do not provide an in-length overview of the article by Hoffmann, Inderst and Ottaviani (2013), as their work is concerned with targeted advertising, in general, and uses political micro-targeting only to demonstrate the generalizability of their framework; however, I do briefly reference to the results derived by the authors.

Further, I provide an overview of empirical studies, which mainly focus on the persuasive capacity of targeted political messaging and discuss behavioural mechanisms driving the observed results. Subsequently, I discuss the implication of the theoretical models in light of the empirical results. In addition, Madsen (2019) provides an extensive overview of data-driven political campaigns in terms of their history, high-level design and behavioural mechanism at play, and I reference his work frequently.

2.1 Micro-targeting and polarization of policy

An article by Prummer (2020) proposes a theoretical model which connects increased media fragmentation in online space to political polarization. The model contains two competing candidates, A and B , a class of attached and a class of unattached voters. Candidates communicate their political platforms via a network of media outlets. Attached voters are informed about the political platforms and will always choose to support their favoured candidate, while unattached voters are poorly informed and can be nudged to change their political preferences if targeted. Compared to traditional mass media outlets such as TV channels, which have relatively small reach, online outlets (e.g. social networks) are open to a large, diverse audience. In terms of the model, social media contain a higher share of unattached voters which makes them a more attractive channel for targeted political advertisement. As is standard in related literature, voters have a quadratic utility of the form:

$$U(c, y|x) = \begin{cases} -(y_c - x)^2 & \text{if } c = A \\ -(y_c - x)^2 + \theta & \text{if } c = B \end{cases} \quad (2.1)$$

Where x is the voter's bliss point (i.e. the peak of his utility function), and partisan preferences θ are distributed according to $\mathcal{U}[-1, 1]$ (c is the index of the candidate). The media are modelled as a bipartite network $\{K, M, N\}$, where K is a set of unattached voters and M is the set of media outlets and $N \subseteq K \times M$ is the adjacency matrix mapping connections of voters to individual outlets. Candidates are solving a problem of choosing the optimal pair $\{x_c, T_c\}$, where x_c is the platform advertised to the targeted voters and T_c is the subset of targeted outlets. Consequently, voters connected to the targeted outlets believe $y_c = x_c$, while non-targeted

voters will stick to their prior beliefs about the platform of c . Intuition already suggests that since the attached candidates vote for their preferred party no matter what information they will get targeting them is much less profitable than targeting unattached voters. To show this formally, Prummer derives a measure of media centrality, which is defined as:

$$W(T) = k(T)(\mathbf{E}(X|T) - \pi_c)^2 \quad (2.2)$$

For some targeted set of media outlets T , with π_c denoting prior beliefs of voters about the c 's platform. Media centrality can be decomposed into two factors $k(T)$, which counts voters within the targeted set and the weighting factor $(\mathbf{E}(X|T) - \pi_c)^2$, which measures the average distance of voters' bliss points from their prior knowledge about candidates platform. The author shows the solution to the candidate's problem, i.e. choosing $\{x_c, T_c\}$ so that the candidate gains maximum support coincides with the maximisation of $W(T)$. In order to connect micro-targeting to polarization, the author recasts her model in a dynamic setting. Voters learn adaptively so that the policy platform of period t becomes the beliefs of the next period $t + 1$.

$$\pi_c^{t+1} = x_c^t \quad (2.3)$$

In the dynamic environment, the platforms offered by the candidates change cyclically; more precisely, the policies of each candidate oscillate between two distinct values $x_{cL} < x_{cR}$ in the long run. The author measures polarisation as an average distance between the long-run policies offered by both candidates.

In addition, the model takes to account the process of media fragmentation, which is essential to its results. Fragmentation of media

is defined as a migration of voters of towards isolated niche media outlets. To give a stylised example, consider a model citizen John who watches the evening news on satellite TV every evening; the channel has an extensive audience all over the country. Recently, John bought a new cellphone and started to use a news app which supplies him with an everyday news briefing based on his selected interests. We might argue that John is better off with his personalised news channel; however, the tradeoff of personalisation is a loss of a common frame of reference with his peers. More formally, if \bar{k}_1 is the average number of voters connected to a media outlet in state 1 and \bar{k}_2 is the average number of voters connected to an outlet in state 2, then fragmentation by the transition from 1 to 2 if $\bar{k}_2 > \bar{k}_1$. A rigorous definition of media fragmentation is provided within the original article.

Formally, there is no resource constraint limiting the choices made by the candidates. However, the distribution of voters across outlets creates an implicit tradeoff. If the candidate wishes to target outlets with high media centrality, he has to skew the offered policy towards the group which dominates its audience, as there are overlaps between audiences of different outlets, which limits the optimal level of policy customisation. With the increased media fragmentation, there are also implicitly fewer overlaps as people tend to gravitate towards the consumption of niche media. As a result, the candidates are able to offer increasingly polarised policies without alienating other voters as a byproduct.

In the context of the model, the efficiency of micro-targeting fundamentally relies on the naive behaviour of the voters. This issue is brought forward by the author; she notes that although the targeting works only towards uninformed voters, the results do not require all voters to be naive since the micro-targeting is a beneficial

strategy as long as there is some remaining share of the electorate uninformed about the policy. The predictions of the model are tested against time series voting data from the House of Representatives. The voting decisions are mapped into one dimension, corresponding roughly to left and right-leaning policies and polarization is computed as a difference of party means. The detrended data clearly display a cyclical behaviour which is in line with the implications of the model in a dynamic setting. Naturally, the observed pattern could be a result of broader socioeconomic influences translating into the political cycle; however, the results seem to be robust with respect to a number of validity checks carried out by the author. Further, Prummer also tests the influence of internet penetration on political polarization; the results of regression analysis show a positive influence of increased internet use on polarization. Again the empirical results appear to be valid against a multitude of validity checks. To sum up, the general implications of the model appear to be in line with the empirical evidence, although given the relatively complicated structure of the model itself, more direct testing would require highly disaggregated data.

Interestingly, a positive effect of internet penetration on opinion polarization is also documented by Sikder et al. (2020), who show that after controlling for an increase in connectivity, internet penetration increases polarisation in opinions. The authors use their findings to complement their network model of social learning, which demonstrates that although increased connectivity between agents causes overall convergence of opinions, it also allows biased individuals to propagate their opinions with greater efficiency, which leads to a persistent polarisation on a societal scale. It is necessary to draw careful distinctions between the polarization of political representation, which is documented by Prummer (2019)

and the findings by Sikdar et al. (2020), which are related to the polarization of individual opinions. The purpose of this comparison is to illustrate that there is a multitude of drivers connected to communication and media technology that can be potentially responsible for the above-described polarization effect. On the other hand, the polarization of the electorate and polarization of political representation are independent and interact in the long run.

2.2 Micro-targeting as a strategic advantage

There is a widespread sentiment that micro-targeting provides political candidates with a competitive advantage in comparison with those who rely on less agile means of advertising. This sentiment is supported by empirical findings which suggest that tailored content, matching the psychological profile of the reader, has indeed enhanced persuasive power. This is documented by several empirical studies, for example, by Dobber et al. (2022) and Tappin et al. (2023).

The model by Titova (2023) attempts to formally identify conditions under which the targeted advertisement provides the political candidate with a significant strategic advantage. In this section, I review the key components of the author's original model in order to illustrate the core mechanics behind the main result. The set-up in Titova (2023) consists of the newcomer candidate who challenges the status-quo incumbent. The key difference is that the status-quo candidate has an established political platform. In contrast, the challenging candidate has uncertain intentions in the eyes of the electorate. The space of admissible policies is confined to an $X = [-1, 1]$, with the status quo normalised to zero. Implicitly the lower and upper bounds represent extremes of the political spec-

trum.

The challenging candidate has a private type $x \in X$, which represents his true policy preferences. As already hinted, the policy of the status-quo candidate is common knowledge. To gain the support of the voters, the challenging candidate sends targeted messages m to the voters, which has a format of subsets of X (i.e. each message corresponds to the interval in $[-1, 1]$). The messages are verifiable, i.e. $x \in m$ for all m . The author assumes that the targeting candidate precisely knows the preferences of the voters; the same assumption is made by Prummer (2019) in the previously reviewed model. The assumption of verifiable information is natural in this context, as it ensures that voters are able to make an inference based on the received messages. Voters have a quadratic utility with a peak v , which is ubiquitous in models involving strategic communication between agents. Beliefs of the voters about the policy platform of the challenging candidate are given by a nondegenerate probability distribution. The payoff of the challenging candidate is normalised so that he receives a payoff equal to 1 in case he wins and 0 otherwise. As already mentioned, the utility of voter i is given by:

$$u_i = -(v_i - x)^2 \tag{2.4}$$

Where v is the bliss point, and x is the policy of the given candidate. Implicitly, the utility of voting for the status quo candidate is constant. Because the policy platform of the status quo candidate is normalised to 0, a net utility of voting for the challenging candidate can be expressed as:

$$\alpha_i(x) = -x^2 + 2vx \tag{2.5}$$

This immediately yields the optimal policy of the voter, i.e. voter

i votes for the challenging candidate if $\alpha_i \geq 0$. The author defines a technical concept of unwinnable elections. The elections are considered unwinnable if, under public disclosure, the challenging candidate has a zero probability of winning. Titova proves that in the baseline model, the challenging candidate can only win if his policy platform coincides with the status quo; however, this is a measure zero event with respect to the voters' beliefs; consequently, there is no strategy with a non-zero probability of winning from the point of view of the challenging candidate if the options are limited to public campaigning. To clarify, public disclosure practically means the impossibility of sending private messages in this context; type x is still private knowledge of the challenger.

The author proves that within this setup, private messaging allows the newcomer candidate to increase his probability of winning, and there is a direct tradeoff between the benefit of the targeting candidate and the welfare of the voters.

Several types of equilibria may arise when private messaging is allowed, to illustrate the basis of mechanisms driving the core results of Titova (2023), I will focus on the candidate-preferred equilibrium, where the challenging candidate maximises his odds of winning the election.

Let us define a few additional objects. Let $A_i = \{x \in X | \alpha_i \geq 0\}$ be an approval set of the voter i , the existence of the approval set is a direct implication of the voter's optimal policy and $W_i^c \subseteq X$ a message sent by the challenging candidate to the voter i . The author defines implementable equilibrium as a situation where the message W_i^c is sufficiently ambiguous (i.e. $A_i \subseteq W_i^c$) but at the same time sufficiently attractive for the voter:

$$\int_{W_i^c} \alpha_i(x) d\mu_0 \geq 0 \tag{2.6}$$

Where μ_0 represents the prior beliefs of the voter about the challenger's policy. Note that in the original model in Titova (2023), the situation is further complicated by the fact the candidate has to simultaneously appease voters on both sides of the political spectrum. So that with the public advertisement, if two voters, i and j , have opposing political preferences (e.g. $v_i = 1$ and $v_j = -1$) and $A_k \subseteq W^c$ must hold for $k = i, j$, it has to be that $W^c = X$. Thus the challenging candidate would be forced to send a completely uninformative message; even though this does not constitute formal proof, it conveys an intuition why the public advertising does not improve the situation of the challenging candidate in the baseline model. Private advertising, i.e. micro-targeting, relaxes this constraint.

Again, suppose there are two different voters. Voter l , with the bliss point v_l on the left side of the political spectrum, that is $v_l < 0$ and voter r with $v_r > 0$. If the micro-targeting is allowed, then the candidate's optimisation problem consists of choosing a pair (W_l, W_r) so that it maximizes his expected payoff given by:

$$\int_{W_l \cap W_r} d\mu_0(x) \quad (2.7)$$

In addition, the solution to the candidate's problem has to satisfy inequality 2.6 for both voters r and l . When the inequality constraint of the i -th voter binds, it means he received the worst policy proposal he is willing to tolerate while still voting for the challenger. With exception of a single case when the approval set of one of the voters is so large the his constraint does not bind, Titova shows that if some pair (W_l^*, W_r^*) solves the candidate's optimisation problem the welfare of the voter is always lower in comparison to the case when the candidate is forced to publicly disclose his campaigning message.

Similar results regarding asymmetric political micro-targeting

are obtained by Hoffmann, Inderst and Ottaviani (2013), who demonstrate that in the case when only one of the candidates uses the strategy, there is a loss of welfare on the side of the voters. Although the underlying model is very different, in both cases, the micro-targeting by only a single candidate leads to suboptimal results from the point of view of the electorate.

2.3 Micro-targeting in empirical research

Because of its tight relation to the issue of online privacy regulation, targeted advertising has rich coverage in empirical research. In general, researchers focus on the persuasive effects of the targeted ads in the political context, their perception by the public or analysis of the information content of targeted messages. To perform content analysis, researchers usually resort to machine learning algorithms. Papakyriakopoulos et al. (2018) analyse the pool of user-generated data on a German social media platform (namely Facebook). This article is particularly interesting since European countries have a much tighter privacy protection framework in comparison with the US, whose data are frequently used to document the practice; this point is also raised by the authors themselves. The strategy is to identify a subset of users who are politically engaged and lack strong partisan preferences. Researchers used Facebook Graph API to mine user data; the result was a data set containing information of more than 1.2 million individuals; the authors clustered identified users according to their partisan preferences. To cluster the posts which were of interest to identified users according to specific topics, a latent Dirichlet allocation was used, which is a machine learning technique from the class of Bayesian hierarchical models.

The purpose of the study was to show that although privacy in Germany is strictly regulated, social media require users to actively agree with the procession of personal data upon signing in to the platform; consequently, the data can be used by third parties as a source of behavioural data for campaign optimisation. The researchers indeed managed to categorise voters according to their political affiliation and create a dataset which could be potentially used as a blueprint for developing a targeted political campaign. Further, the authors outlined potential targeting strategies, showing that the stringent regulation is of little use, unless it takes into account a wider context; they argue that users usually opt in voluntarily and have to accept an extensive list of terms and conditions regarding data processing, which is not necessarily informative, as the scope and broader implications of such an agreement are not clearly communicated to the consenting individual.

The observation that the majority of data which serve as a base for developing models for targeted advertisement is supplied on a voluntary basis by the users themselves is an example of a more general phenomenon known as the privacy paradox. Privacy paradox is a term used to describe the discrepancy between the self-declared preferences for privacy and the actual willingness to invest resources into privacy protection (Uno, Sonoda, Bessho, 2021). Naturally, the existence of a privacy paradox hinders the efficacy of the current regulation model, which largely shifts the burden of responsibility on the individual.

Similarly to Papakyriakopoulos et al. (2018), Ortega (2021) runs an analysis of topics on the selected dataset of targeted political ads in various high-stakes election campaigns in Europe. The author hypothesizes that because targeted advertisement reduces the risk of spillovers outside of the target audience, the content of target-

ed advertising should be consequently more diverse in comparison with en masse published ads. The results of the statistical analysis support the proposition; however, the identified differences are relatively weak. A possible explanation of the weak result lies in the sampling method used to collect the underlying data. The dataset is provided by a nonprofit organization and was collected by individuals on a voluntary basis through a plugin installed in their web browser. Consequently, the sample variance is most likely significantly lower in comparison with the general population.

Dobber et al. (2022) measured the efficacy of personalised political advertisement during the course of the real election in the Netherlands; the authors found a positive influence of personalised messages on the likelihood of voting for the targeted political party. It is necessary to point out that the research sample was small and with highly skewed educational attainment, as 92% of the participants had some form of higher education. This is a striking disparity in comparison with the general population. In combination with a small sample size (124 participants), the results are somewhat hard to generalise.

A series of experiments conducted by Tappin et al. (2023) compares the persuasive power of targeted advertising with two other benchmark strategies commonly used in political campaigns. Namely, the authors compared the performance of a naive campaigning strategy, which uses a set of preidentified policy topics and assigns the target audience at random; a single best message strategy which supplies a single high-impact narrative to a diverse audience and micro-targeted messaging, which diversifies advertised topics across individuals based on their individual profiles. The authors report that the micro-targeted campaign outperformed the remaining benchmark strategies by a margin of 70% on average, depending

on context. Interestingly, much of the persuasive power of targeted advertising can be attributed to a single covariate, i.e. for a successful campaign; it is sufficient to divide the audience based on a single characteristic. This largely undermines the widespread sentiment that a successful targeted campaign requires extensive data coverage.

Regarding the indirect effects of micro-targeting on political competition, the survey by Matthes et al. (2022) identifies a negative impact on overall trust in a democratic process; authors further note that perceived micro-targeting increases interest in political engagement in the test subjects. Lavigne (2020) documents the positive effect of targeting on voters' partisan preferences for their favoured political party.

2.4 Discussion

The study of the implications of targeted political advertising is a relatively niche topic; however, due to the important policy implication, especially with regard to the online privacy regulation design, the area is extensively covered by empirical research across multiple behavioural disciplines. This thesis presents only a selection of related empirical literature, with emphasis being put on the most recent works.

That being said, empirical research in this area faces multiple challenges, which make it difficult to draw some final conclusions. Most importantly, there is a lack of complete data to run analysis on, as the information about the design of political campaigns is by its nature non-public; the researchers have to rely on data collected by volunteers, which are generally biased in comparison with the overall population. The alternative is to conduct direct

surveys; however, this confines the analysis to small samples and makes it hard to capture systemic impacts. Several studies tackle this issue by mining data directly from social media platforms. While this approach allows a bulk topic and sentiment analysis of political content and identifies user engagement, it does not permit researchers to monitor how is the advertised content distributed across different audiences during the campaign phase. Despite the methodological difficulties described, there appears to be a general consensus that micro-targeted advertising has enhanced persuasive power in comparison with less sophisticated schemes of content distribution (Tappin et al. (2023), Dobber et al. (2022)). Further, there is a bulk of indirect evidence that people are more responsive towards personalised messages which match their personal values and appeal to their individual experience (Madsen, 2019).

The theoretical research in this area faces even further difficulties. Firstly, as the diversity of results in empirical studies implies, the success of targeted campaigns realises through multiple behavioural and social pathways. Consequently, it is problematic to pinpoint a common denominator which would serve as a starting point for the construction of a rigorous theoretical model. Generally, the lack of spillovers is identified as the main systematic advantage the personalised advertisement provides. This point is explicitly made, e.g. by Papakyriakopoulos et al. (2018). Both reviewed models, that is, Prummer (2019) and Titova (2023), exploit this stylised fact in their work.

In Prummer (2019), the increased media fragmentation of online outlets increases candidates' returns on diversifying their respective political platforms, which in turn leads to a higher polarization over time. The increased returns on targeted advertising are directly caused by the reduction of spillovers, i.e. the messages are not

observed by voters outside of the intended audience. The differences in the topology of communication networks provided by traditional mass and that of online social platforms appear to be crucial for understanding the causal pathways from the theoretical point of view. However, the analysis of network systems is difficult without certain simplifying assumptions (e.g. the assumption of k -regular graphs used in the network literature). Prummer (2019) notes that without assuming a regular media fragmentation process, there is generally not a monotonic connection between the fragmentation of media and the polarization of policy. Further, as online platforms allow active engagement of users among each other, not only the mode of communication between voters and candidates but also the subsequent transmission of political information among users themselves could have a significant influence on campaign success. This can be illustrated by the briefly referenced model by Sikdar et al. (2020), who study the influence of social networks on opinion polarization. An additional complication is introduced by the endogeneity of communication networks (Bolletta, Pin, 2022). Similar indirect effects are not native to traditional means of advertising, where the information can spread only through the immediate social circle of the receiver. However, the online environment allows communication between arbitrary participants and social circles are consequently much more extended in comparison to real-life social networks, which tend to be spatially restricted.

The work of Titova (2023) focuses on the differences between public and private messaging and does not address the differences in the structure of communication networks. Both models imply increased diversity of micro-targeted messages in contrast with campaigning aimed at the broad public. This hypothesis is explicitly tested by Ortega (2021), who identifies only a weak increase in diver-

sity in targeted ads. Regarding the persuasive efficacy of the targeted ads, both models assume that advertisement by the candidates is informative by default. Consequently, the persuasive effects are embedded into the models by assumption. Prummer (2020) shows it is optimal to target cross-pressured voters whose prior beliefs about the candidate’s platform are far from their favoured policy outcome; this is more or less in line with the empirical evidence. Although targeting voters with preexisting partisan preferences is also beneficial as it strengthens party affiliation (Lavigne, 2020).

Overall, it is difficult to take make a definite value judgment regarding net on the societal impact of targeted political campaigns. Some sources take a strong normative position on the issue; for example, the article by Burkell and Regan (2019) presents an extensive survey regarding its supposed undesirable externalities. The authors argue that the increasing algorithmization of search engines promotes passive confirmation bias and creates an environment in which individuals lack common grounds for critical engagement with targeted political advertisements, which in turn leads to the reinforcement of preexisting political biases. The effects of passive confirmation bias caused by algorithmic customisation of search content based on the user’s browser history are also referenced by Sikder et al. (2020) within their network model of opinion polarization. Thus similar sentiment is not uncommon in literature and appears to be also backed by indirect evidence stemming from behavioural research (Madsen, 2019); however, it is hard to reconcile with relatively mild effects observed during empirical studies.

However, it could be the case that the persuasive effects of the targeted campaigns are in the natural context (i.e. outside of an experimental setting) increased by the reduced exposure to opposing opinions, which is not present in other types of media to such an

extent.

Although many traditional media outlets have a preexisting bias with regard to certain political topics and cater to a specific audience, they generally cannot accommodate the worldview of a specific reader. In contrast, online platforms, for various reasons, actively adjust the content presented to the user based on his preferences revealed by past engagement with similar types of content. Consequently, the targeted advertising of political topics does not have to present a problem on its own, but in the broader context, it feeds into the overall tendency of personalised search optimisation to lock users in an informational bubble.

Further concerns regarding the practice mentioned, for example, the consolidation of political competition. Due to their high cost, similar approaches are generally more likely to be used by larger political parties which have access to the necessary funding, leading to increased consolidation of political power. Further, if the voters receive heterogeneous information about the political platform of the same candidate, the deliberation process becomes less efficient as the voters lack a common frame of reference. However, although there is a significant amount of empirical proof regarding the persuasive power of political marketing, the intensity of these effects is largely unknown; it is unlikely that targeted advertisement triggers a significant change in political preferences. The interpretation that it helps cross-pressured voters to discover alignments between their existing preferences and the political platform of targeting the candidate seems much more plausible in light of the reviewed research. The main objective issue associated with targeted political advertising appears to be the potential misuse of private data, as the legal framework is often ambiguous and does not clearly identify the responsible party (Blasi, Vermeulen, 2021). Thus legal frame-

work should outline detailed guidelines in this regard, and the campaigning process should be transparent. In addition, voters should be aware when they are they are viewing customised political content, which would further aid transparency and prevented potential misinformation. Additionally, messaging containing extremist sentiment, which increases tension between different socioeconomic groups, poses a certain security risk (Briant, 2017).

To conclude this section, targeted political marketing appears to increase the efficiency of political campaigns in a significant manner. It allows a broad coverage, reduces spillovers and takes advantage of automated machine learning algorithms to customise the advertised content. However, provided they follow carefully outlined legal and ethical frameworks, the concerns regarding negative social externalities appear to be low in general; despite the overwhelmingly negative sentiment of some studies, the measurable impacts appear to be statistically significant, but it is unlikely to cause any systematic disbalance in the political competition. Although, if there are significant differences in the campaigning strategies, the micro-targeting candidate will likely have a competitive advantage. In the European context, the main problem appears to be a heavy emphasis on individual responsibility in the area of privacy regulation. The existence of the privacy paradox, which is repeatedly encountered in the context of online privacy protection, suggests that the efficacy of such a framework is low; in the context where the broader implications of data processing agreements are often unclear as there is increasing informational asymmetry between users and technological companies providing online services.

3. Model

3.1 Introduction to the model

In this section, I discuss the overall structure of the model. The aim is to provide a concise overview of its key components. In addition, I also hint main results of the model and the mechanisms behind them. The following section formalises the primitives of the model, such as the types of agents, their initial beliefs, available actions and model timing.

The model illustrates a public choice problem with asymmetric information. Two political candidates, A and B , compete to gain public support for their respective agenda. At the end of the game, a final policy decision is implemented, which can take on an arbitrary value from the $[-1, 1]$ interval. Each candidate $c \in \{A, B\}$ has private political beliefs given by type θ_c , which constitutes a bliss point of his concave utility function and determines his preferences over the set of policy outcomes, while the θ_c is a private knowledge of the candidate I further assume that both political candidates are well informed and can accurately guess the type of their opponent ex-ante. Consequently, θ_c is only unknown to voters. The type θ_c can take on values from the set $\Theta = \{-1, 0, 1\}$. The candidates with $\theta_c = 0$ hold a moderate political position, while those with $\theta_c = \pm 1$ hold extreme views on the opposing sides of the political spectrum. The electorate is a continuum of measure zero voters denoted by N and indexed by i . The set of voters is normalised so that N is of measure one as a whole. Similarly to candidates, each voter has a single-peaked concave preference regarding candidates' political platforms with a bliss point given by his respective private type $\theta_i \in \Theta$. The voters engage in political discourse on a shared

online platform, where they can make public statements regarding their position towards future policy. I assume that participation in the discourse is inherently enjoyable to the voters; however, there are trade-offs associated with public expression stemming from the fact that the voters also care about the opinion of their peers when expressing their views. The voters who do not wish to reveal their political preferences are referred to as silent. Further, each voter leaves a trace of private information which is only observable by the provider of the online social platform and identifies the voter's private type. Strictly speaking, the term political preferences refers to the agent's ranking of available political platforms or ranking of policy outcomes; however, within the text, I use the term to refer to the type θ if no confusion can arise.

The candidates attempt to gain support from the electorate with a mixture of public and private campaign promises. To optimise their campaign strategy, candidates can utilise voters' private information; however, utilising private data is costly compared to using directly observable information from public sources. After the elections, each candidate gains legislative power in proportion to his respective share of votes. Consequently, the end policy has a format $\omega_A\pi_A + (1 - \omega_A)\pi_B$, where ω_A is the share of the electorate voting for A and $(1 - \omega_A)$ is the share of voters who support B . I assume that a minimum amount of commitment is required, i.e. candidates cannot push for a policy they did not advertise during their campaign; furthermore, the candidates can only send private messages which correspond to their true political agenda. I posit such an assumption is not restrictive since, without any form of commitment, the campaign promises would be uninformative for the voters; thus, the forced restriction of the candidate's actions does not change the final result. However, I do not support this

conjecture by formal proof. The assumption is reminiscent of the restriction used by Titova (2023), who allows political candidates to signal only policy proposals which do not principally contradict their true agenda.

The game is played sequentially; at the initial stage $t = 0$, the public discussion unravels, and voters signal their opinions about possible policy solutions by either supporting some political position from the set $\{-1, 0, 1\}$ or they remain silent and do not show public support for any option in Θ . As already mentioned, the voters take into account possible adverse implications of backing up their favoured political position when choosing their public signalling action. Following stage $t = 1$ the candidates observe the results of the public debate and choose their campaigning strategy, which consists from a public policy proposal π_c^P and subsets of the electorate T_c targeted by the given candidate c . In order to optimise their campaign strategy, the candidate can strike a deal with the service provider and purchase private data about the voters; the use of private data comes with an additional cost proportional to the size of the surveyed population. At the last stage of the model, $t = 3$, the voters cast a private vote to support their favourite candidate. As already mentioned, the end policy is a convex combination of the agenda pushed by each respective candidate with weights corresponding to their respective share of votes. The model is supposed to illustrate the inherent difference between political advertising via mass media, where the interaction is one-sided, and the advertised message gets simply delivered to the end consumer and targeted advertising on social media which allows for information transmission in both directions. In contrast to the traditional mode of media communication, social media allow lively interactions in both ways and, more importantly, they allow each user to observe

actions taken by his peers. Consequently, concerns about one’s reputation may skew the public discourse, as observed, for example, by Burnett, Knighton and Wilson (2022). Because the data about the private content engagement are ultimately logged in by the service providers, they are not effectively private but can be purchased for marketing purposes. It is only a matter of strength of incentive if political actors take advantage of this option. If the public discourse is self-censored, the candidates with extreme agenda are more likely to be interested in the private data of users since the voters with extreme opinions apply a higher rate of self-censorship.

Further, with sophisticated voters, the presence of targeted political advertisements decreases the informational value of policy proposals publicly announced by candidates; thus, voters outside of the targeted audience cannot efficiently discriminate between different candidates based on their public proposals.

3.2 Primitives of the model

If possible, the primitives of the model are set in line with existing frameworks in related literature. More specifically, I use the quadratic utility, which is ubiquitous in the literature on strategic communication and literature concerned with voting behaviour. The set of admissible policies is set to be the $[-1, 1]$ interval on the real line, where the endpoints represent polar opposites of the political spectrum. Similar abstractions are used in the reviewed works of Prummer (2020) and Titova (2023).

3.2.1 Voters

Types and preferences

The set of voters is denoted by N . Individual voters are of measure zero and indexed by i and N is of measure one (i.e. $\int_N di = 1$). Voters have private types θ , which express their ideological stance towards the proposed policies. The ideological type θ takes on values from a set $\Theta = \{-1, 0, 1\}$ and each value occurs with an equal probability of $1/3$. Voters with $\theta = 0$ are moderate, while those with $\theta = \pm 1$ hold strong opinions on the opposite sides of the political spectrum. In addition, each voter has a type α which determines his baseline utility from participation in the online public discourse. Type α is independently distributed across voters according to $\mathcal{U}[\alpha_l, \alpha_h]$ and similarly to type θ it is a private knowledge of each voter. Both types are independent of each other.

Each voter has a utility function u which allows him to form preferences over the political platforms of the individual candidates. Because the true value of the political platform is the private knowledge of each candidate, from the perspective of the voter, the platforms are random variables. For some voter, i , the platform of candidate c is a random variable distributed according to i 's beliefs. The utility function of the voter i with an ideological type $\theta_i \in \Theta$ has the following form:

$$u_i(x_c|\theta_i) = -(x_c - \theta_i)^2 \quad (3.1)$$

Where x_c belongs to the support of the political platform of the candidate c . Let $\mathcal{X}^i = \{X_c^i\}_{c \in M}$ be a sequence of political platforms of the individual candidates distributed according to the beliefs of

some voter i and define \mathcal{X}_i^* as:

$$\mathcal{X}_i^* = \arg \max_{X_c^i \in \mathcal{X}^i} \{\mathbf{E}_i[u_i(x_c)]\} \quad (3.2)$$

The sub-sequence \mathcal{X}_i^* represents political platforms favoured by the voter i (to clarify $\mathcal{X}_i^* = \mathcal{X}_{c=c^*}^i$, similarly $X_i^* = X_{c=c^*}^i$). I assume that in the end the voter always has to support only a single alternative. Therefore if there is only a single element $X_i^* \in \mathcal{X}_i^*$ the voter will choose to support X_i^* . If the maximum in 3.2 is not unique the voter will choose his preferred alternative by randomising over the elements of \mathcal{X}_i^* with equal probabilities, leading to a unique choice of X_i^* .

As already hinted, the voters also derive an intrinsic utility from publicly expressing support for their preferred political platform. Any voter i can signal their support for the political platform $X_i^* \in \mathcal{X}^i$ by sending a public signal $s_i(X^*) = 1$ and receive a utility $v_i(s_i(X^*) = 1)$.

$$v_i(s_i(X^*)) = \begin{cases} \alpha_i + \beta \int_{N-i} \mathbf{E}_i u_j(X_i^*) dj & \text{if } s_i(X^*) = 1 \\ 0 & \text{if } s_i(X^*) = 0 \end{cases} \quad (3.3)$$

The interpretation is following, each voter can receive a fixed utility equal to his type α_i by publicly signalling his support for his favourite policy proposal X_i^* (and consequently his support for the proponent of X_i^*). However, depending on the overall level of conformity given by the value of β , the voter also considers the opinion of his peers on X_i^* . In case the disutility from the possible public backlash is greater than the α , the voter chooses to keep his opinion private and does not support X_i^* (hence $s_i(X_i^*) = 0$ for all $X_c^i \in \mathcal{X}_i^*$). To clarify, $v_i(s_i(X_c^i = 1)) < 0$ for all $X_c^i \neq X_i^*$ and

$v_i(s_i(X_c^i = 0)) = 0$, i.e. it is never optimal to support a platform not in line with one's ideology. The intuitive interpretation of this condition is that lying by commission is not allowed. For a private signalling action, $s(X_c^i)$, the agent's utility is given by:

$$v_i(s_i(X^*)) = \begin{cases} \alpha_i & \text{if } s_i(X^*) = 1 \\ 0 & \text{if } s_i(X^*) = 0 \end{cases} \quad (3.4)$$

The total utility U_i of the voter i received at the end of the game is given by the sum of the total value of his signalling actions and the utility from the end policy π implemented after the elections. So that:

$$U_i = u_i(\pi|\theta_i) + \sum_t v_i(s_i^t) \quad (3.5)$$

Where s_i^t is the signalling action taken by voter i in time $t = 0, 1, 2$. The signalling action of $t = 1$ is void since the voter does not move at stage $t = 1$ as specified in the following section.

Actions available to voters

Actions available to the voter differ through out the stages of the game. However, they all have a similar structure. In stage t of the game, each voter i is provided with a sequence of political platforms, $\mathcal{X}_t^i = \{X_{k,t}^i\}_{k \in I}$, where I is the indexing set. Further, the upper index i signifies that the beliefs about platforms $\mathcal{X}_{k,t}^i$ may potentially differ across voters. The voter can take either private or public signalling action expressing his approval for his preferred policy $X_i^* \in \mathcal{X}_t^i$. The formation of preferences over the elements of \mathcal{X}_t^i is described in the section 3.2.1 above. The signalling action has a format of a sequence $\{s_i(X_{k,t}^i)\}_{k \in I}$, i.e. sequence indexed by the set I , this is to ensure that upon observing $\{s_i(X_{k,t}^i)\}$ the receptor of

the message can precisely identify if the voter i supports some platform $X_{k,t}^i$. To give an example, suppose that in the second stage ($t = 2$) of the game, two candidates, A and B , publicly propose policies x^A and x^B , respectively, and there is no additional uncertainty. Because the proposal is public, the observed sequence of platforms $\mathcal{X}_2 = (X_{A,t=2} = x^A, X_{B,t=2} = x^B)$ is common to all voters. Voter i prefers $X_{A,t=2}$ and wishes to express public support; hence $X_i^* = X_{t=2}(A)$ and the signalling action of voter s_i will take the form of $s_i = (s_i(X_{A,t=2}) = 1, s_i(X_{B,t=2}) = 0)$. Voter j also prefers proposal $X_{A,t=2}$ (hence $X_j^* = X_{A,t=2}$); however, he chooses to keep his stance private so that $s_j = (s_j(X_{A,t=2}) = 0, s_j(X_{B,t=2}) = 0)$.

As already discussed, each voter $i \in N$ has his preferred platform X_i^* which is unique. Implicitly the signalling actions of the voters always contain at most one non-zero entry. In case no ambiguity can arise, the excessive notation can be discarded, and we can write only $s_i = (1, 0)$ and $s_j = (0, 0)$. The following paragraphs contain a detailed description of the actions available to the voters.

Let A_t^i be the set of actions available to voter i at stage t and $S_{t=0}^i$ the corresponding strategy.

$t = 0$: In the initial stage, the voters will receive a sequence of platforms $\mathcal{X}_0 = (x_0^1 = -1, x_0^2 = 0, x_0^3 = 1)$, announced by nature. In other words, in the initial phase of the game, the voters receive a policy proposal corresponding to each element in the set of ideological types Θ . The voter will respond by a public signalling action $s_i \in \{0, 1\}^{|\Theta|}$ and will receive the corresponding utility $v_i(s_i)$. Thus in stage $t = 0$, the set of actions available to arbitrary voter i is $A_{t=0}^i = \{(s_i)\}$ and his strategy $S_{t=0}^i$ is a map such that:

$$S_{t=0}^i : \{\theta_i, \alpha_i, \mathcal{X}_0\} \longrightarrow (s_i) \quad (3.6)$$

The voter's strategy will have a similar structure throughout the whole game and is explicitly stated to avoid ambiguity.

$t = 1$: The voters do not move at $t = 1$. Thus $A_{t=1}^i = \{\emptyset\}$.

$t = 2$: In the third stage of the game, each voter will receive signals from the candidates informing him about their political platforms. Based on the received signals, the voters will update their beliefs about the platforms of individual candidates. Hence the sequence of platforms $\mathcal{X}_{t=2}^i = \{X_{c,t=2}^i\}_{c \in M}$ is a sequence of random variables indexed by the set M . Each voter then chooses private signalling action s_i to vote for his preferred candidate. Thus, the set of actions available to arbitrary voter i is $A_{t=2}^i = \{(s_i)\}$. Finally, the strategy of i is defined as:

$$S_{t=2}^i : \{\theta_i, \mathcal{X}_2^i\} \longrightarrow (s_i) \quad (3.7)$$

Prior beliefs of the voters

Each voter i has a prior belief about the political platform of candidate c . Voters update their beliefs based on the observable actions taken by other agents. Prior to receiving any signal, the voter's beliefs about the platform of candidate c are given by distribution $\mathcal{B}_i^0(x_c)$ with the following properties:

$$\mathcal{B}_i^0(x_c) = \begin{cases} 1/3 & \text{if } x^c = -1 \\ 1/3 & \text{if } x^c = 0 \\ 1/3 & \text{if } x^c = 1 \end{cases} \quad (3.8)$$

Voters are assumed to be Bayesian.

3.2.2 Candidates

Preferences of candidates

There are two competing candidates, A and B ; the set of candidates is denoted by M . Similarly to voters, candidates have a private type $\theta \in \Theta$ which describes their political orientation. Unlike the voters, who form preferences over the political platforms of individual candidates, the candidates have preferences over the set of possible policy outcomes. Thus if Z is a set of all possible outcomes, then preferences of candidate $c \in M$ can be expressed through a utility function u_c .

$$u_c(z|\theta_c) = -(z - \theta_c)^2 - b_c, \quad z \in Z \quad (3.9)$$

Where b_c is the campaign spending of candidate c .

Cost of private information

Actions involving publicly available information are considered to be costless in the model. Any candidate can strike a deal with a third-party data provider and use the de facto private data about voters to identify their political preferences. In the real-world context, the data provider is typically the owner of the social media platform. In addition, there are companies which specialise in the tracking and collection of user data for marketing purposes. In the context of the model, this third party is considered to be politically neutral, and its objectives do not explicitly enter in. The candidates can choose to utilise the private information of silent voters who did not reveal their types in the public discourse. If a candidate decides to purchase private information about voters in some set $H_c \subset N$, he

will pay a cost b_c proportional to the volume of the set H_c .

$$b_c = \phi \int_{H_c} di \quad (3.10)$$

Where ϕ is assumed to be a non-negative constant expressing the cost of surveying an individual unit, for simplicity, I suppose that ϕ directly corresponds to the degree of regulatory privacy protection. In general, the ϕ would be an increasing function of multiple additional factors such as technology and degree of market competition. However, for the sake of argument, I suppose that the technology for information transmission is nearly perfect, and the additional costs of private information are dictated purely by the degree of regulatory restrictions associated with their use.

There are several subtle points worth addressing before moving forward. Firstly, there is a distinction between privacy as understood in the legal context and privacy in the model. As the thesis later moves to policy interpretation of the model results, this is necessary to clarify. In the legal context, information such as names, locations and user-generated content associated with social media accounts could be considered private; factually, this is not the case; depending on the platform, at least a subset of listed data can be publicly accessed. This is demonstrated, e.g. by Papakyriakopoulos et al. (2018), who mine user-level data through the public API from German Facebook sites. This is possible since users opt in on a voluntary basis and give their formal permission by approving the terms and conditions agreement. Thus although private in theory, such data are accessible at low cost and consequently de facto public. A country can thus have stringent regulations on the handling of private information. However, if the burden of responsibility rests disproportionately on the individual, the level of effective regulation is still low.

Relating this back to parameter ϕ , the high value of ϕ corresponds to stringent regulation *and liability on the side of the data provider*. This is in line with economic intuition since if the liability rests on the user; there is no additional cost for the company handling data which were de facto provided on a voluntary basis.

Actions available to candidates

Let A_t^c be a set of actions available to arbitrary candidate $c \in M$ at period t and S_t^c corresponding strategy.

$$\mathbf{t=0:} \quad A_0^c = \emptyset$$

$\mathbf{t=1:}$ At $t = 1$, candidates update their beliefs about the types of individual voters based on the observed signalling actions. The candidate's campaign strategy consists of public policy proposal π_c^P , which can take on values $\{\theta_c, 0\}$, a set of targeted voters $T_c \subset N$ and a set of voters whose private information is being purchased H_c . The $H_c \in \{\emptyset, O\}$, where $O \subset N$ is the set of voters who had chosen not to support any policy platform at $t = 0$. The choice of targeted policy proposal signalled to voters in T_c is fixed to $\pi_c^T = \theta_c$. Hence $A_1^c = \{(\pi_c^P, T_c, H_c)\}$.

$$S_1^c : \{\{\mathcal{B}_c(\theta_i)\}_{i \in N}, \theta_c\} \rightarrow (\pi_c^P, T_c, H_c) \quad (3.11)$$

Where $\{\mathcal{B}_c(\theta_i)\}_{i \in N}$ are beliefs of candidate c regarding the ideological preferences of individual voters.

$\mathbf{t=2:}$ At $t = 2$ each candidate has the option to implement his preferred policy π_c in proportion to the support he received from the electorate. The only constraint on the choice of π_c is that $\pi_c \in \{\pi_c^P, \pi_c^T\}$, i.e candidates cannot push agenda which they did not

initially propose in some form. Thus $A_2^c = \{(\pi_c)\}$.

$$S_2^c : \{\theta_c\} \rightarrow \pi_c \quad (3.12)$$

Prior beliefs of the candidates

Prior to receiving any information, each candidate c has belief $\mathcal{B}_c^0(\theta_i)$ about the type θ of an arbitrary agent i .

$$\mathcal{B}_c^0(\theta_i) = \begin{cases} 1/3 & \text{if } \theta_i = -1 \\ 1/3 & \text{if } \theta_i = 0 \\ 1/3 & \text{if } \theta_i = 1 \end{cases} \quad (3.13)$$

3.3 Solution of the model

The solution concept is the perfect Bayesian equilibrium.

3.3.1 Optimal signalling actions of the voters

Proposition 1 *Given the sequence of proposals \mathcal{X}_0 , voters with type $\theta = 0$ will support $X_0 = 0$ with probability p . The voters with types $\theta = \pm 1$ will support proposals $X_0 = \pm 1$ respectively, with a probability q . In any equilibrium $p \geq q$.*

Proof: Because individual voters are of measure zero, they do not take into account the impact of their signalling actions on the end policy π , in other words:

$$\max_{s_i^0} U(\pi, s_i^0, s_i^2) = u_i(\pi) + v_i(s_i^2) + \max_{s_i^0} v_i(s_i^0) \quad (3.14)$$

Consider a voter $i \in N$, with $\theta_i = 0$. If $\mathcal{X}_0^i = \mathcal{X}_0$, then $X_i^* = 0$

and consequently:

$$v_i(s^0) = \begin{cases} \alpha_i + \beta \int_{N_{-i}} u_j(0) dj & \text{if } s^0 = 1 \\ 0 & \text{if } s^0 = 0 \end{cases}$$

The optimal policy is to choose $s^0 = 1$ if $v_i(s^0 = 1) > 0$ and $s^0 = 0$ otherwise. Consequently, we have to evaluate the following probability:

$$\mathbf{P}(\alpha_i + \beta \int_{N_{-i}} u_j(0) dj > 0 | \theta_i = 0)$$

Which is equivalent to:

$$\mathbf{P}(\alpha_i > \beta(2/3)) = 1 - \frac{\beta(2/3) - \alpha_l}{\alpha_h - \alpha_l} = p$$

For a voter j with $\theta_j = -1$ the probability:

$$\begin{aligned} \mathbf{P}(\alpha_j + \beta \int_{N_{-j}} u_l(-1) dl > 0 | \theta_j = -1) &= \mathbf{P}(\alpha_j - \beta(5/3) > 0) = \\ &= 1 - \mathbf{P}(\alpha_j < \beta(5/3)) = 1 - \frac{\beta(5/3) - \alpha_l}{\alpha_h - \alpha_l} = q \end{aligned}$$

Set $p - q = \delta$ then:

$$\delta = \beta(5/3 - 2/3) / (\alpha_h - \alpha_l) = \beta / (\alpha_h - \alpha_l) > 0 \quad \square$$

For computational convenience, I fix $\alpha_l = 0$ and $\alpha_h = 5/3$. Consequently $\beta(5/3)$ always belongs to the $[0, 5/3]$ interval for all $\beta \in [0, 1]$ and p can be expressed as:

$$p = q + \beta \tag{3.15}$$

3.3.2 Posterior beliefs of the candidates

At $t = 0$, nature provides an initial sequence of political platforms $\mathcal{X}_0 = (-1, 0, 1)$. Because initially there is no uncertainty, each voter would prefer to support the platform corresponding to the focal point of his utility function u . **Proposition 1** states that:

$$\mathbf{P}(s_i^x = 1 | \theta_i = x) \begin{cases} p & \text{if } \theta_i = 0 \\ q & \text{if } \theta_i = \pm 1 \end{cases} \quad (3.16)$$

With $p > q$. Assume that voter k chooses not to support any of the platforms in \mathcal{X}_0 , i.e. $s_k = (0, 0, 0)$. Denote $(0, 0, 0)$ as o , the posterior beliefs of an arbitrary candidate c about the voter k who signals $s_k = o$ are:

$$\mathcal{B}_c(\theta_k | s_k = o) = \begin{cases} \frac{1-q}{3-p-2q} & \text{if } \theta_k = -1 \\ \frac{1-p}{3-p-2q} & \text{if } \theta_k = 0 \\ \frac{1-q}{3-p-2q} & \text{if } \theta_k = 1 \end{cases} \quad (3.17)$$

For voters who choose to support any of the platforms in \mathcal{X}_0 , there is no posterior uncertainty about their respective types θ . For the purpose of further analysis, I will refer to the set of silent voters as O .

$$O = \{k \in N : s_k = o\} \quad (3.18)$$

3.3.3 Posterior beliefs of the voters

Public proposals from the candidates are partially verifiable; that is, candidate c with type θ_c can either signal public $\pi_c^P = \theta_c$ or $\pi_c^P = 0$. Consequently, proposal $\pi_c^P = 0$ has an ambiguous informative value as the posterior belief of the voter depends on the best responses on the candidates in equilibrium. However, the signal $\pi_c^P = \theta_c$

is already verifiable because only a candidate with a given type could choose such a proposal, thus after observing $\pi_c^P = \theta_c$ the beliefs of voter i are equal to $\mathbf{P}_i(\theta = \theta_c | \pi_c^P = \theta_c) = 1$. Private messages are completely verifiable; thus targeted voters always know the type of candidate who is targeting them. If the voter belongs to the set of targeted voters T_c of some candidate c he has perfect knowledge about the type θ_c of the candidate in question; thus, each candidate has only incentive to target voters with the same political preferences as his.

3.3.4 Optimal campaign strategy of the candidate

Proposition 2 *In equilibrium, both candidates A and B will choose to publicly propose a moderate policy, independently of their type θ_c . Assuming $\theta_B \neq \theta_A$.*

Proof: Suppose that choices (T_A, H_A) and (T_B, H_B) are fixed. Let $\omega_c(x, y)$ be a share of voters obtained by a candidate c , conditional on $\pi_A^P = x$ and $\pi_B^P = y$.

$$\omega_c(x, y) = \int_{T_c} di + \int_{F_c(x, y) \setminus T_c} di + \frac{1}{2} \int_{G(x, y)} di \quad (3.19)$$

$$(3.20)$$

Where $G(x, y)$ is the set of non-targeted voters such that:

$$G = \{i \in N : \mathbf{E}_i u_i(X_A | \pi_A^P = x) = \mathbf{E}_i u_i(X_B | \pi_B^P = y)\} \quad (3.21)$$

And $F_c(x, y)$ is the set of voters who strictly prefer c to the opposing candidate. Suppose the action profile is such that both candidates signal $\pi^P = 0$ and the candidate with $\theta_c \neq 0$ deviates by signalling $\pi_c^p = \theta_c$. Because $\mathbf{P}(\theta = \theta_c | \pi_c^p = \theta_c) = 1$ for all $c \in M$, thus upon observing any policy proposal, such that $\pi_c^p = \theta_c$ from c , there is no

ambiguity about c 's political platform. Thus voters with $\theta_i = \pi_c^P$ will back up the deviating candidate c ; however, the remaining $2/3$ of the electorate will support his opponent. Consequently, the term $\frac{1}{2} \int_{G(x,y)} di$ will be equal to zero. In comparison, the volume of the set $F_{c'}$ of the opposing candidate c' will increase. In contrast, when signalling c signals moderate policy, the opposing candidate will gain at most $1/2$ share of the electorate, which is a strictly better outcome for c . Thus choosing $\pi_c^P = 0$ is the optimal strategy for any candidate regardless of his type. \square

Proposition 3 *Public policy proposals are uninformative in equilibrium.*

Proof: It follows directly from **Proposition 2** that all types are equally likely to choose $\pi_c^P = 0$ as their optimal campaigning strategy. Consequently, upon observing $\pi_c^P = 0$, arbitrary voter $i \in N \setminus (T_A \cup T_B)$ has a posterior belief $\mathbf{P}_i(\theta_c = x | \pi_c^P = 0)$ about any candidate $c \in M$ and it holds that :

$$\mathbf{P}_i(\theta_A = x | \pi_A^P = 0) = \mathbf{P}_i(\theta_B = x | \pi_B^P = 0) \quad (3.22)$$

For all $\theta \in \Theta$. \square

3.3.5 A moderate candidate running against an extreme candidate

Assume $\theta_A = -1$ and $\theta_B = 0$. Clearly, the case with $\theta_A = 1$ is symmetric to a case with $\theta_A = -1$, and it is sufficient only to analyse one of them.

The strategy of extreme candidate

Assuming that candidate B is best responding to A 's action. It is optimal for A to choose $\pi_A^P = 0$ and target voters with $\theta = -1$; consequently, the only remaining degree of freedom in A 's strategy is the choice of H_A . Let a_c be a course of action taken by candidate c , assuming that B targets all voters with $\theta = 0$; the payoffs of A can be written as:

$$u_A((\pi_A^P = 0, H_A = \emptyset, T_A^*), a_B) = -(1/2 + 5/3\beta)^2 \quad (3.23)$$

$$u_A((\pi_A^P = 0, H_A = O, T_A^*), a_B) = -1/4 - 4/3\phi\beta \quad (3.24)$$

Consequently, A will choose to break the privacy of voters in O if the following holds:

$$\phi \leq 3 \frac{(1/2 + 5/3\beta)^2 - 1/4}{4\beta} \quad (3.25)$$

Denote the value for which the expression above holds with equality by ϕ_A . Because p and q are functions of β , we can express ϕ_A as:

$$\phi_A(\beta) = 3 \frac{(1/2 + 5/3\beta)^2 - 1/4}{4\beta} \quad (3.26)$$

Clearly, $\phi_A(\beta)$ is non-negative for any $\beta \geq 0$, i.e. the existence of some positive price unit price ϕ_A below which it is a profitable strategy to collect private data of voters is guaranteed. A similar condition can be obtained for voter B ; again, I assume that opposing candidate A targets all voters with $\theta = -1$.

$$u_B((\pi_B^P = 0, H_B = \emptyset, T_B^*), a_A) = -1/36(3 + 2/3\beta)^2 \quad (3.27)$$

$$u_B((\pi_B^P = 0, H_B = O, T_B^*), a_A) = -1/4 - 4/3\phi\beta \quad (3.28)$$

The threshold price for candidate B is given by:

$$\phi_B(\beta) = 3 \frac{1/36(3 + 2/3\beta)^2 - 1/4}{4\beta} \quad (3.29)$$

Similarly to ϕ_A , the ϕ_B is always positive for $\beta \geq 0$, however $\phi_A > \phi_B$ for all admissible values of β . Consequently, the set $\Phi = [\phi_B, \phi_A]$ is always nonempty. If $\phi = \phi'$ such that $\phi' \in \Phi$ candidate A will choose to use the private information of the voters and will gain proportional support equal to $\omega_A = 1/2 + 1/9\beta$ share of the electorate. Clearly, three distinct scenarios can arise.

Weak privacy protection ($\phi \in [0, \phi_B]$): In the scenario when the cost of using private data is low, it is beneficial for both candidates to enhance their campaigning strategy by the use of private data because it is beneficial regardless of the strategy of the opponent. Consequently, the advantage stemming from the utilisation of private data is diminished in equilibrium. While both candidates would be better strictly off sticking to the use of public data, however, the deviation from $H_c = O$ can be exploited by the opponent, and consequently, the candidates are stuck in the socially undesirable equilibrium. Thus the low level of privacy protection in the presence of micro-targeting turns election campaigning into a prisoner's dilemma. Because privacy is not hardwired into the utility of voters, there is no direct loss of utility from surveillance.

Medium privacy protection ($\phi \in [\phi_B, \phi_A]$): As already hinted, in the intermediate case, the moderate candidate is effectively outpriced from the private data market. Implicitly, the extreme candidate is able to gain majority support and skew the end policy π to his liking. Extreme candidate, A is supported by a $1/2 + 1/9\beta$ share of voters; thus, increasing individual conformity combined with insufficient

privacy protection empowers the proponents of extreme policies. Let ΔW_N be an aggregate change of voters' utility resulting from the change of end policy and $N_{\theta'} \subset N$ be a community of voters with type $\theta = \theta'$. In the case of low privacy protection, the resulting policy is $\pi_l = -1/2$, while in the current case of medium privacy protection, the resulting policy is $\pi_m = -(1/2 + 1/9\beta)$ thus:

$$\Delta W_N = \int_N [u_i(\pi_m) - u_i(\pi_l)] di = \Delta W_{N_{-1}} + \Delta W_{N_0} + \Delta W_{N_1} \quad (3.30)$$

For the subset of voters with $\theta = -1$ the change in total utility is given by:

$$\Delta W_{N_{-1}} = \int_{N_{-1}} [u_i(\pi_m) - u_i(\pi_l)] di = -(1/2 - 1/9\beta)^2 + 1/4 > 0 \quad (3.31)$$

For moderate voters with $\theta = 0$ the change is:

$$\Delta W_{N_0} = \int_{N_0} [u_i(\pi_l) - u_i(\pi_m)] di = -(1/2 + 1/9\beta)^2 + 1/4 < 0 \quad (3.32)$$

And for extreme voters with type $\theta = 1$:

$$\Delta W_{N_1} = \int_{N_1} [u_i(\pi_l) - u_i(\pi_m)] di = -(3/2 + 1/9\beta)^2 + 9/4 < 0 \quad (3.33)$$

Consequently, the total change in welfare is negative and can be expressed as:

$$\Delta W_N = -(1/2 + 9/2\beta + 2/9\beta^2) \quad (3.34)$$

In comparison with the low levels of privacy, if $\phi \in [\phi_B, \phi_A]$, there is a welfare loss for the moderate segment of the electorate as the policy becomes more extreme; in addition, extreme voters on the

opposite side of the political spectrum lose welfare due to the fact that new policy π_m is relatively further from their focal points. The loss of welfare by extreme voters with type $\theta = 1$ is partially balanced by the gain of voters with $\theta = -1$; however, due to the concavity of the quadratic utility function, there is still a net loss as well. Counter-intuitively there is thus a range of ϕ such that it increases the equilibrium support of the extreme candidate. This would imply that if there is privacy protection, it has to be strictly enforced; otherwise, there are undesirable consequences. Naturally, if the β is close to zero and the public debate is very open regarding the diversity of opinions, this effect is diminished.

High level of privacy protection ($\phi > \phi_A$): If the ϕ is sufficiently high, the candidates are unwilling to use private information regardless of the strategy of their opponent. Further, the support of the moderate candidate increases by a factor $(1 - q)/6$, which yields a policy π_h given by:

$$\pi_h = \frac{\beta - 2}{6} \tag{3.35}$$

Thus under the strict protection of privacy, the resulting policy is more moderate than in the previous case leading to overall welfare improvement of voters.

3.4 Discussion

The results show that if voters put nonzero weight on the utility of their peers (i.e. $\beta > 0$), then there are three regimes of campaigning competition between moderate and extreme candidates, which are dictated by the exogenously set cost of private data. As I assume that such costs have a direct correspondence to the degree

of regulatory stringency on the use of private data, the results can be interpreted in a policy context. With the low cost of private information, there is an overwhelming incentive for both moderate and extreme candidates to track silent voters through the use of their private data. Because in the equilibrium, both candidates gain equal support for the policy outcome, there is no difference in the policy outcome. In contrast, if the regulatory requirements are such that ϕ belongs to $[\phi_B, \phi_A]$ interval, the moderate candidate has a handicap in comparison to the extremist because the probability that A with type $\theta_A = -1$ will find sufficient support in the group of silent voters is simply much higher than it is for the moderate candidate B . This result has potential regulatory implications towards setting repercussions for breach of individual privacy. It implies that insufficient or ambiguous regulatory rules might create an asymmetry between fringe political entities and their more moderate counterparts. In addition, the presence of targeted advertising raises uncertainty about public political proposals, as they carry little informational weight from the point of view of the non-targeted groups of voters.

A sufficient increase of ϕ beyond the threshold ϕ_A devaluates gains of the extreme candidate A and brings the policy closer to the mean preference of the voting population resulting in utility gains as the utility of voters with $\theta = 1$ increases at a higher rate than the utility of voters with opposing type. Restricting targeting, in general, would lead to further welfare improvement as candidate A would be forced to publicly signal a policy platform directly corresponding to his type θ_A .

The main distinction between the hereby presented model and existing literature is the consideration of the process by which the candidates gather data about the political preferences of the elec-

torate. The notion that the data which are directly accessible by the general public might be biased in comparison with actual unobservable political preferences is not new; note on this issue is made, for example, by Papakyriakopoulos et al. (2018), who use the data from public API in their study. Here, I assume that public data are skewed in a systematic manner due to the existence of self-censorship in the online political discourse. This assumption is based on anecdotal evidence as well as on empirical evidence provided by Burnett, Knighton, Wilson (2022) and Madsen, Verhoeven (2016), who show that people filter publicly shared information if they are concerned about their public reputation. Given that political leanings usually are an exceptionally sensitive topic, the motivation to avoid a bad reputation can be strong. In the context of my model, the described effect leads to public discourse being biased in favour of moderate opinions. As a result, moderate candidates know their target audience. In contrast, the extreme candidate needs additional data to identify potential supporters and is motivated to put up with additional costs.

In contrast to the original thesis proposal, I assume that all candidates have uniform access to technology and financial resources. From a purely pragmatic standpoint, micro-targeted campaigns are much more likely to be carried out by larger political parties that have access to necessary financial and technological resources. But this is universally true, and it is not specific to a particular type of political campaign.

It is true that newly formed political parties or candidates entering political competition would likely have higher returns on targeted advertising as their platform is novel and unknown to the electorate; however, this effect is already addressed in depth by Titova (2023). Within the framework of my model, both candidates enjoy

the benefit of prior uncertainty about their platforms.

There are several possible extensions to the current version of the model. Firstly, the set of types Θ could coincide with the set of admissible policies; this would allow for a greater diversity of political positions. Similarly, the set of signalling actions accessible to them could be extended to the $[-1, 1]$; in combination with the verifiability condition proposed by Titova (2023), the candidates would still be able to make credible election promises. I posit that under the current specification, where voters have identical prior beliefs about both candidates, the extension of available actions would be to the advantage of the moderate candidate as he could target voters on both sides of the political spectrum. Additionally, the process which enforces conformity in the public discourse could be explicitly modelled instead of direct enforcement through the utility function of the voter.

To conclude this section, strong regulation of privacy appears to have a positive effect in the discussed context. Furthermore, for regulation to yield the desired result, the legal responsibility has to rest on the correct subject. Placing the burden of the responsibility on the data provider increases compliance costs which in turn raises the cost of private data for the political candidate, yielding socially desirable outcomes. From a pragmatic standpoint, this appears as a plausible solution since the data provider typically has the necessary technological capacity to ensure the safety of the data as well as the necessary knowledge, while the individual user mostly lacks such resources. Similar asymmetry exists, for example, in the financial industry, where consumers of investment products often enter into risky contracts which they are they are not fully qualified to understand. And the alignment of the objective of the banker and consumer is ensured by a complex regulatory framework. Addition-

ally, the consumer has limited liability for his individual choices as some risks (such as back default) are covered by mandatory insurance. Similarly, contracts about the collection of private data are often too complex for consumers to assess implied risks accurately; thus, the legal framework should be calibrated with the same caution. Under current conditions, this is not the case, as the users can often provide implicit agreement by simply entering a website.

4. Conclusion

There is a multitude of joint empirical and theoretical evidence that micro-targeting is not neutral in the political context but can significantly influence policy outcomes and attitudes of the voters. By its nature, targeted campaigning requires, in some capacity use of individual-level data and consequently increases the risk of infringement of personal privacy. From the practical point of view, the fact that the target political advertisement incentivises hazardous treatment of private user data is well documented by the case of the Cambridge Analytica scandal, which is universally viewed as an unethical and socially undesirable outcome by a broad public.

Whether privacy as such is a valuable commodity on its own is an ongoing debate within modern economic literature, the survey article by Acquisti, Taylor and Wagman (2015) shows that answer to this question varies greatly depending on the context. In the commercial setting, targeting individual customers might decrease their search costs and improve the seller's profit which is a mutually desirable outcome. In contrast, in the political setting, the welfare effects on voters are largely unclear, with recent literature indicating that has rather undesirable systemic side effects for voters.

The model by Hoffmann, Inderst and Ottaviani (2013) shows that if only one of the candidates uses micro-targeting, there is a decrease in voter welfare. More recent work by Prummer (2020) indicates a connection between micro-targeting and polarization of policy in the context of the increasing fragmentation of online media outlets. In addition, Prummer (2020) shows that micro-targeting reduces the welfare of non-targeted voters; however, the author does not discuss the implications for the welfare of voters in detail. Last but not least, the recent model by Titova (2023) implies that tar-

geted advertising aids the challenging candidate to maximise their chance of winning the elections in case they run against an incumbent with an established political platform. However, the challenging candidate improves his odds of winning by decreasing the welfare of voters, which is a socially undesirable outcome.

In this thesis, I presented a minimum case model which attempts to take into account incomplete information on the side of the candidates. I assume that the use of private information is costly for the candidates in comparison with the use of public data; the additional costs are a result of privacy protection regulations and enter as an exogenous parameter.

My analysis shows that if the information contained in the public data is skewed in favour of the consensus opinion, then the extreme candidate has a higher incentive to boost his campaigning strategy by exploiting the private data of voters. I support the assumption about behavioural conformity on social media with direct evidence from two empirical studies (Burnett, Knighton, Wilson, (2022) and Madsen, Verhoeven, (2016)). A direct result of such asymmetry between the extreme and moderate candidates is the increased political influence of the extreme candidate. This results in a welfare loss of the voters with different policy preferences. Such an effect is conditional on the cost of private information being set sufficiently high.

Because within my model, the costs of private information are synonymous with the degree of regulatory privacy protection, my results can be interpreted in the policy context. The model can exist in three distinct regimes. The weak regulation of privacy encourages candidates to collect private information regardless of their type. As I explicitly avoid assuming that there is any inherent utility to privacy, the utility of voters is not affected. This would naturally

change in the case when voters care about the breach of privacy itself. If the mode of regulation is such that the costs of utilising private information discourage the moderate candidate but not the extremist, the resulting policy is skewed in favour of the extreme candidate in comparison with the preceding case. Consequently, there is a loss of welfare for moderate and opposing voters, which results in an overall decrease in welfare in general.

The high cost of private information (i.e. strict privacy protection) discourages both candidates from entertaining the use of private data. As a result, the policy implemented after the elections is more moderate, as it is easier for a moderate candidate to identify his supporters based solely on public information, and there is a welfare improvement for moderate and opposing voters in comparison with both preceding cases. Thus it can be concluded that strict privacy protection is the welfare-maximising choice in the context of my model. Putting this into the context of previously reviewed theoretical research, there is an indication that targeted political advertising has adverse effects on the welfare of voters. In addition, Prummer (2020) also identifies possible long-run effects on the polarisation of policy.

Reviewed empirical research identifies several common themes with respect to political micro-targeting. Firstly, targeting seems to increase an individual's likelihood of voting for a given candidate; this is documented, for example, by Dobber et al. (2022) as well as by Tappin et al. (2023). In general, the empirical literature explains the persuasive effects of targeted advertisements through the increased compatibility between the message and psychological characteristics of the targeted individual (Madsen, 2019). Asymmetric use of micro-targeting by only one of the candidates could thus skew election results in his favour. Such findings support the

conclusions of hereby received theoretical works. According to a study by Matthes et al. (2022), the awareness of targeted advertising decreased trust in democracy in surveyed individuals; however, it also increased the overall interest in political engagement.

There are additional valid concerns regarding the undesirable side effects of the targeted political campaigns. For example, Burkell and Regan (2019) argue that due to the tendency of algorithmic searches to promote content which is a priori in line with individual biases, the targeted political campaigns take advantage of the increasing polarisation of opinions and further cement the overall divergence of individual politics. Although such claims can be supported by extensive indirect evidence, the results from direct empirical studies suggest rather mild effects on individual opinions. However, further research in this regard is necessary.

In conclusion, there is a significant amount of evidence that political micro-targeting can influence election results. Within my thesis, I argue that even if the negative effect of privacy breach is dropped from the utility function, the use of private data in a political context still has undesirable effects for the majority of voters in terms of policy outcomes. Thus it is not generally desirable to allow political candidates to diversify messaging based on private data. Similar conclusions can be drawn from other theoretical models which are concerned with this issue. Although they do not address the problem of privacy protection directly, all the above-described negative effects rely on the ability of the candidate to partition voters on the individual level. On the flip side, micro-targeting can promote awareness about political issues or increase voter turnout, but it seems safe to conclude that its use by political parties or other distinct interest groups should be, at a minimum, carefully monitored and subject to strict legal and ethical boundaries.

Additionally, the regulatory framework for online privacy needs to reflect informational asymmetries between technological companies providing online services as schemes of processing private data become increasingly difficult to navigate for an average user.

5. References

5.1 Academic references

1. Acquisti, A., Taylor, C. R., & Wagman, L. (2015). The Economics of Privacy. *SSRN Electronic Journal*.
<https://doi.org/10.2139/ssrn.2580411>
2. Blasi, C., & Vermeulen, M. (2021). Reflections on the murky legal practices of political micro-targeting from a GDPR perspective. *International Data Privacy Law*, 11(4), 348–359.
<https://doi.org/10.1093/idpl/ipab018>
3. Bolletta, U., & Pin, P. (2022). Polarization When People Choose Their Peers. *SSRN Electronic Journal*.
<https://doi.org/10.2139/ssrn.4067169>
4. Burkell, J., & Regan, P. M. (2019). Voter preferences, voter manipulation, voter analytics: policy options for less surveillance and more autonomy. *Internet Policy Review*, 8(4).
<https://doi.org/10.14763/2019.4.1438>
5. Burnett, A., Knighton, D., & Wilson, C. (2022). The Self-Censoring Majority: How Political Identity and Ideology Impacts Willingness to Self-Censor and Fear of Isolation in the United States. *Social Media and Society*, 8(3), 205630512211230.
<https://doi.org/10.1177/20563051221123031>
6. Dobber, T., Trilling, D., Helberger, N., & de Vreese, C. (2022). Effects of an issue-based microtargeting campaign: A small-scale field experiment in a multi-party setting. *The Information Society*, 39(1), 35–44.
<https://doi.org/10.1080/01972243.2022.2134240>

7. Galeotti, A., Ghiglino, C., & Squintani, F. (2013). Strategic information transmission networks. *Journal of Economic Theory*, 148(5), 1751–1769. <https://doi.org/10.1016/j.jet.2013.04.016>
8. Hoffmann, F., Inderst, R., & Ottaviani, M. (2013). Hypertargeting, Limited Attention, and Privacy: Implications for Marketing and Campaigning. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2395208>
9. Jann, O., & Schottmüller, C. (2019). An Informational Theory of Privacy. *The Economic Journal*, 130(625), 93–124. <https://doi.org/10.1093/ej/uez045>
10. Lavigne, M. (2020). Strengthening ties: The influence of microtargeting on partisan attitudes and the vote. *Party Politics*, 27(5), 965–976. <https://doi.org/10.1177/1354068820918387>
11. Madsen, J. K. (2019). *The Psychology of Micro-Targeted Election Campaigns*. <https://doi.org/10.1007/978-3-030-22145-4>
12. Madsen, V. T., & Verhoeven, J. W. M. (2016). Self-censorship on Internal Social Media: A Case Study of Coworker Communication Behavior in a Danish Bank. *International Journal of Strategic Communication*, 10(5), 387–409. <https://doi.org/10.1080/1553118x.2016.1220010>
13. Matthes, J., Hirsch, M., Stubenvoll, M., Binder, A., Kruike-meier, S., Lecheler, S., & Otto, L. (2022). Understanding the democratic role of perceived online political micro-targeting: longitudinal effects on trust in democracy and political interest. *Journal of Information Technology & Politics*, 19(4), 435–448. <https://doi.org/10.1080/19331681.2021.2016542>

14. Ortega, L., A. (2021). Are microtargeted campaign messages more negative and diverse? An analysis of Facebook Ads in European election campaigns. *European Political Science*, 21(3), 335–358. <https://doi.org/10.1057/s41304-021-00346-6>
15. Papakyriakopoulos, O., Hegelich, S., Shahrezaye, M., & Serrano, J. C. M. (2018). Social media and microtargeting: Political data processing and the consequences for Germany. *Big Data and Society*, 5(2), 205395171881184. <https://doi.org/10.1177/2053951718811844>
16. Prummer, A. (2020). Micro-targeting and polarization. *Journal of Public Economics*, 188, 104210. <https://doi.org/10.1016/j.jpubeco.2020.104210>
17. Castanho Silva, B., & Proksch, S. O. (2021). Politicians unleashed? Political communication on Twitter and in parliament in Western Europe. *Political Science Research and Methods*, 10(4), 776–792. <https://doi.org/10.1017/psrm.2021.36>
18. Sikder, O., Smith, R. E., Vivo, P., & Livan, G. (2020). A minimalistic model of bias, polarization and misinformation in social networks. *Scientific Reports*, 10(1). <https://doi.org/10.1038/s41598-020-62085-w>
19. Tappin, B. M., Wittenberg, C., Hewitt, L. B., Berinsky, A. J., & Rand, D. G. (2023). Quantifying the potential persuasive returns to political microtargeting. *Proceedings of the National Academy of Sciences*, 120(25). <https://doi.org/10.1073/pnas.2216261120>
20. Titova, M. (2023). Targeted advertising in elections. Accessible on the author’s website: <https://maria-titova.com/research/>

21. Ullah, I., Borelli, R., & Kanhere, S. S. (2022). Privacy in targeted advertising on mobile devices: a survey. *International Journal of Information Security*, 22(3), 647–678.
<https://doi.org/10.1007/s10207-022-00655-x>
22. Yosuke Uno & Akira Sonoda & Masaki Bessho (2021). The Economics of Privacy: A Primer Especially for Policymakers, *Bank of Japan Working Paper Series*, 21-E-11, Bank of Japan.

5.2 Non-academic references

1. Briant, L.E. (2017). *Evidence for the US Senate Judiciary Committee on Cambridge Analytica and SCL Group*. US Senate Judiciary Committee
2. Statista (2023) Digital Advertising. *Market Insights report*.
<https://www.statista.com/outlook/dmo/digital-advertising/worldwide>